DOCUMENT RESUME

ED 459 769 IR 058 263

AUTHOR	Tillett, Barbara B.
TITLE	A Virtual International Authority File.
PUB DATE	2001-08-00
NOTE	6p.; In: Libraries and Librarians: Making a Difference in the Knowledge Age. Council and General Conference: Conference Programme and Proceedings (67th, Boston, MA, August 16-25, 2001); see IR 058 199.
AVAILABLE FROM	For full text: http://www.ifla.org.
PUB TYPE	Opinion Papers (120) -- Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Authority Control (Information); Bibliographic Records; Cataloging; *Databases; International Programs; *Internet; Models; *Shared Resources and Services

ABSTRACT

A pool of authority records for bibliographic entities (persons, corporate bodies, works/expressions, concepts, objects, events, and places) to use on the Internet is of interest, not only to libraries and their users, but also to publishers, copyright and rights management organizations, museums, and archives. This paper explores different models of how such a file might work. Authority control remains the most expensive part of cataloging, but through cooperative efforts like NACO (Name Authority Cooperative), SACO (Subject Authority Cooperative), and IFLA (International Federation of Library Associations and Institutions) initiatives, the research done in one library can be shared internationally to lower the cost. (Author/MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 459 769

67th IFLA Council and General Conference August 16-25, 2001

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

A.L. Van Wesemael

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Code Number: 094-152a-E
Division Number: IV
Professional Group: Cataloguing
Joint Meeting with: -
Meeting Number: 152a
Simultaneous Interpretation: -

A Virtual International Authority File

Barbara B. Tillett
Director, Integrated Library System Program Office
Washington, DC USA

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Abstract:

A pool of authority records for bibliographic entities (persons, corporate bodies, works/expressions, concepts objects, events, and places) to use on the Internet is of interest not only to libraries and their users but also to publishers, copyright and rights management organizations, museums, and archives. This paper will explore how this all might actually work. Authority control remains the most expensive part of cataloging, but through cooperative efforts like NACO, SACO, and IFLA initiatives, the research done in one library can be shared internationally to lower the cost.

It has often been observed that the current Web is chaotic for finding information. It needs help and we can provide it! Introducing an element of authority control to the Web environment would help meet these objectives:

- to facilitate the sharing of the workload to reduce cataloguing costs - and our community has expanded, especially in Europe these days, where libraries are viewed with archives, museums, and rights management agencies as "memory institutions." Shared authority information will reduce costs overall. Other objectives for authority control are
- to simplify the creation and maintenance of authority records internationally and
- to enable users to access information in the language, scripts, and form they prefer or that their local library provides for them.

The virtues of authority control have been debated and restated for decades. When we apply authority control in the Web environment, we are reminded how it brings precision to searches, how the syndetic structure of references enables navigation and provides explanations for variations and inconsistencies, how the controlled forms of names and titles and subjects help collocate works in displays, how we can

IR058263

BEST COPY AVAILABLE

actually link to the authorized forms of names, titles, and subject that are used in various tools, like directories, biographies, abstracting and indexing services, and so on…We can use the linking capability to include library catalogues in the mix of various tools that are available on the Web. Controlling forms used for access and displays provides consistency for users.

We are all aware of very poor OPACs that lack cross references or links to authority files and without these features, quite frankly, they are not catalogues!

Over the past few years there have been several projects that help us get closer to providing authority control on a global scale. In my paper[1] for the LC Bicentennial Conference on Bibliographic Control in the New Millennium last November, I described the AUTHOR Project (European Union), the report of the IFLA MLAR (Minimal Level Authority Records) that identified essential data elements needed in authority records (today we'd call these metadata); the IFLA FRANAR (Functional Requirements for Authority Numbers and Records (you will have just heard about this from Françoise Bourdon); Dublin Core/Authorities; discussions about CORC authority records for a possible global expansion to build an authority file; more global capability with multi-scripting with Unicode in Windows; and the expansion of NACO and SACO to AACR2 and LCSH (Library of Congress Subject Headings) to users worldwide. I won't go into more detail today, but my paper also indicated the recent focus on the need for interoperability and how we can now map different communication formats with Z39.50 protocols (in fact the LEAF Project explores this model).

There are also crosswalks to the "MARCs" including XML, ONIX, and others. These crosswalks can help us to search and retrieve library resources effectively with abstracting and indexing services and other resources on the Web. All of these technological capabilities are coming together now and we are really at the brink of making a virtual international authority file a reality.

We are also making an historic change to how we view Universal Bibliographic Control. The IFLA UBC principles for authority control are parallel to those for bibliographic control, namely that 1) each country is responsible for the authorized headings for its own personal and corporate authors (they didn't mention uniform titles, series, or subjects), and 2) the authority records created by each national bibliographic agency would be available to all other countries needing authority records for those same authors.

In the 1960's and 1970's when this was really catching on, technology had not yet advanced to make such sharing practical on an international level. Plus the lack of funding for an international center to manage such a program prevented that visionary concept from becoming reality.

For the past couple of years a new view of Universal Bibliographic Control is emerging from several working groups within IFLA. This new perspective reinforces the importance of authority control, yet puts the user first…It's a practical approach that recognizes a user in China may not want to see the heading for Confucius in a Latinized form or a strange pinyin romanized form, but in their own script.

Yet to still get the benefits of shared authority work and creation of bibliographic records that can be re-used worldwide, we can link authorized forms of names, titles, and even subjects through the authority files of national bibliographic agencies and other regional agencies to create a virtual international authority file. There are several models for how this might work and we need to do more pilot projects of prototypes of models to test which would be best to pursue.

In order to be of most use to the library users in each country, the scripts should be the scripts they can read! What a novel idea! Transliteration may serve as a way for some users to be able to decipher records, but the accuracy of using original scripts is much better. We should now provide at least cross references for variant forms of headings in variant scripts when that is appropriate. In the United States MARBI is starting to explore this possibility and more work needs to be done. We should eventually be able to display the script and form of a heading that the user expects and wants.

I believe that many of us realize the value of parallel authority records for the same entity that allow us to set up the syndetic structure of cross references and authorized forms of headings to be used in catalogues intended for a specific audience and include variants in alternate scripts at least as cross references.

As we look at linking we must recognize that different cataloguing rules have differences in what are considered entities - AACR2's choices are not universal, for example, German rules do not recognize that the ships logs can be under an entry for the name of the ship, so they would not have an authority record for ships names, but AACR rules would have such an entry. Similarly for events, for meetings of corporate bodies, the German rules would not create a heading for the entity that AACR2 creates in as a hierarchically subordinate heading for Meeting under the name of the corporate body.

There are also different practices for undifferentiated names - the Germans recently changed their rules to differentiate more names, but they used undifferentiated forms for personal names using just initials for forenames. However, we all know that even under the same cataloguing rules, say AACR2, we find we can make a new separate record for an entity when we get more information to differentiate that person, so the record for the undifferentiated name can reflect different associated entities over time.

So how do we get to a virtual international authority file? Several major authority files exist, built according to their own cataloguing rules and rule interpretations. It would be helpful to conduct a one-time project to link the existing records for the same entity - a retrospective matching project. One suggestion has been to use matching algorithms, such as those developed by Ed O'Neill and others at OCLC, building on bibliographic clues for machine matching at a fairly high level of accuracy. We would still have manual matching and checking to do, but could get a long way with machine matching. We could also have the computer add the linking text strings and record control numbers to facilitate later links and pathways to preferred forms for displays.

Some local systems already provide us with computer-assisted mechanisms for automatic checking of headings against an existing authority file, and we could see this expanded to then launch a search against a virtual international authority file, if no match was found locally.

We can also envision the capability of displaying the found matches from the virtual file for a cataloguer to edit or to merge information, if desired, into the local authority record, including capturing the information for future linking.

Some systems now provide community specific retrievals to concentrate on the subject needs of a community in selecting resources for online searches, and other systems like "my library" or "my opac" even go beyond that to individual specific retrievals. Those could build in the authority preferences for user preferred scripts and displays for controlled vocabularies.

We want to have the authorized form preferred by a library as the default offered to most users, but we can also envision offering user-selected preferences through client software, or "cookies" that let the user specify once what their preferred language, script, or cultural preference is - for example for spelling preferences when cultures have variations, like American English and spelling preferences in the United Kingdom - labor and labour...

Other ways to do this that have been suggested over the years are standard numbers, like the ISAN, INSAN, ISADN, etc., but I would prefer that we test using the unique, persistent record control numbers and see if that works. That saves having to set up another international organization to manage the distribution and maintenance of such numbers. In my paper for the LC Bicentennial Conference, I provided several scenarios of how this might play out. Let's quickly take a look at two.

A cataloguer types in information into a bibliographic record and the local system checks the local authority file. The local system finds the record in the local authority file and displays it so the cataloguer can confirm it is the same entity. Then, we'd like the system to automatically update the bibliographic record with the authorized information from that authority record, once confirmed by the cataloguer.

Now what about the scenario of no record in the local file? A cataloguer types in information, and the local system checks the local authority file. The response is there is no match, so the system tells the cataloguer and allows the cataloguer to launch a Web search for the virtual international authority file. With the Internet running much faster than it does now, we would quickly get the results, in this case, the match with a record created at the National Library of Russia in St. Petersburg (see figures). Our cataloguer takes a look and perhaps does not want all the information but likes a reference or two and wants a link. The local system asks the cataloguer if she wants the system to create a basic authority record from the one found and to make a link to it…and we say yes... Our local system then builds a local authority record, grabbing the linking information from the virtual authority file - that is the record from St. Petersburg. The cataloger then adds the 100 authorized form according to the locally used cataloguing rules, in this case AACR2, and our cataloguer can add other fields if needed. The local system adds the linking 700 field - the MARC format has the 7xx fields in authority records, where we can put the linking authorized form and the record control number and the source information for future linking. This linking of authority files would primarily be among the national or regional authority files of national bibliographic agencies - depending on the model we choose. I'll come back to that in a minute.

So we've now added another link in the virtual international authority file to the authorized form following AACR2 and the Russian record for the same entity following the Russian cataloguing rules in Cyrillic script. Our local system updates our local bibliographic record.

When a user comes along, the local system or the cookies on the user's system, could specify the user wants to see the Cyrillic form, and we could display it for them. You can also imagine displaying any script or a Braille keyboard output, or we could provide voice recognition response, built on a user's profile or their "cookie."

Some of the possible models for a virtual international authority system are: a distributed system that searches the independent authority files of national bibliographic agencies using a Z39.50 search; a linked system using a Z39.50 search on linked authority records (this increases the precision of the search); a centralized system such as using the Open Archive Initiative (OAI) model that creates a union authority file of metadata on server linked to national authority files; or another model is a central link system where all of the authority files from all of the national bibliographic agencies are linked to one.

With the distributed and linked Z39.50 models, a user could key in a term that perhaps that only appears as a cross reference in one file so it is retrieved, but also linked to another record in another file, so that too is retrieved for that same entity. We may find that this model is the best approach in terms of record maintenance.

The Open Archives Initiative model creates a server with harvested metadata from the national authority files, and that information is refreshed in the server whenever there are changes in the national files. This means the day to day record maintenance activities continue to be managed as they are now by the national bibliographic agency (or regional authority). Unless we also build in the linking, we possibly will lose a level of precision in the searching in this model; but there are ways to include the links for entities in this model, too. There are many variations of models we could imagine.

Another model would be using one central authority file and link all others to that, so that work would not need to be done by each national bibliographic agency with all other participants in this international universe. A cataloguer would then get access to all the authority records for that entity worldwide by a single search of the central file. If there was not match in that central file, a search could then be made with Z39.50 to the other files.

I am sure you can imagine other variations of these models. And we need to try them out to see which will be best for us in today's Internet environment.

I offered the following four recommendations in my paper:

1) test this concept of linked authority files;

2) establish a Z39.50 profile for authority records (this is underway as an extension of the Bath Profile);

3) work with local system vendors to enhance this capability of local systems enabling the creation of authority records and searching the Web for resource authority records; as well as the other half from the user's perspective to display their preferred script or orthography; and

4) work with creators of search engines on the Web to expand this concept to databases and resources they search and bring some control to the chaos.

This gives you some ideas of how cataloguers can build authority records on the Web and then, once the authority control structure is there worldwide, this can include other stakeholders (publishers, rights management agencies, archives, museums, and other libraries) - all can use this information and reduce costs. Authority control will help users of the Web to benefit from collocation and search precision that authority control enables and do it in ways that are meaningful to users in their preferred language and script.

---

[1] Tillett, Barbara B. "Authority Control on the Web," Bicentennial Conference on Bibliographic Control for the New Millennium, Nov. 2000. Available on the Web as (with underline before word paper) http://lcweb.loc.gov/catdir/bibcontrol/tillett_paper.html